

DOCUMENT RESUME

ED 171 788

ED 009 262

AUTHOR McGuire, Christine
 TITLE Issues in the Assessment of Continuing Professional Competence.
 PUB DATE Apr 79
 NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)
 EDRS PRICE MF01/PC01 Plus Postage.
 DESCRIPTORS *Certification; *Evaluation Methods; *Evaluation Needs; Health Personnel; Job Skills; Medical Care Evaluation; *Personnel Evaluation; Professional Continuing Education; Professional Occupations; *Professional Personnel; Standards; Task Performance; Work Sample Tests
 IDENTIFIERS *Competence

ABSTRACT

Methods of assessing professional competence involve both licensing procedures, as well as assessment based on satisfactory performance of tasks associated with a professional's duties. Relicensing and/or mandated continuing education courses should not be used to ascertain professional competence in place of a demonstration of acquired skills because continuing education programs are not necessarily quality controlled; may not be appropriate measures of performance; and do not take into account such factors as situational constraints and pressures. Job performance should be adequately sampled in natural situations or in carefully structured simulations. Questions raised by determining competence via performance are: (1) what professions and what individuals within each profession should be assessed and how often?; (2) what is the most effective cycle for reassessment?; (3) what range or types of competencies should be evaluated?; (4) what evaluation techniques are available?; (5) what criteria should be used in judging performance?; and (6) how should standards for continued licensure or certification be set? (MH)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

Christine McGuire

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

ISSUES IN THE ASSESSMENT OF CONTINUING PROFESSIONAL COMPETENCE*

Christine McGuire**

The Nature of a Profession

What is it that defines a profession and what, if anything, distinguishes professions from other occupational and vocational groupings? Psychologists and sociologists who have studied this question have enumerated many social-psychological factors involved in the professionalization process. However, the characteristics of special interest to the discussion which follows are those concerned with monitoring the competency of members of a profession.

In this context Professor Matarazzo (1977) has noted in a recent paper that with the exception of theology, each of the other learned professions seems to have proceeded through roughly the following stages of evolution: first an era of random and haphazard entry into the field by individual practitioners with an interest in practicing the profession; second a loose organization of these practitioners into a voluntary guild or professional association that helps both to give identity to the profession and to facilitate intraprofessional communication among its members; third, slowly evolving but increasingly more visible and stringent entry requirements for membership, followed by equally visible and stringent definitions of appropriate curricula for training; fourth, more formal self-policing of these educational curricula via accreditation by national review groups chosen by the profession itself; fifth, certification by some agency outside the profession of the minimal competency of graduates of these programs and their licensure to practice; followed, sixth, by additional certification of higher levels of competence through specialty or related boards within the profession; seventh, subsequent establishment of requirements for periodic relicensure verifying current minimal competency issued through the same governmental machinery that awarded the initial license; and ultimately, eighth, periodic review by local peer groups to assure that the individual practitioner has maintained at least minimal competency, and recertification of higher or specialty level competency by periodic practice audits, self-assessments or reexamination through the guild's own specialty board. Matarazzo concludes that "the profession of medicine has been a model for other professions in highlighting both the advantages and disadvantages that can accrue from these various levels of internal and external controls on the profession."

*Invited Address Division I, AERA. Annual Meeting. San Francisco. April 9, 1979.

**Associate Director for Research and Development. Center for Educational Development. University of Illinois at the Medical Center. Chicago, IL.

The Philosophic Basis for Competency Assessment

In these days of increasingly vitriolic attacks on organized medicine, and of more and more exaggerated malpractice claims against individual practitioners, I am sure that members of the profession will be happy to know that medicine is a model for something. However, the generalized description of the evolution of a profession quoted above probably exaggerates the conceptually neat distinction between initial licensure of minimal competency, specialty certification of greater competency, and peer assessment of continued competency.

Actually, there are basically only two approaches to assessing professional competence, whether initial or continuing. In one approach members of a profession are defined in terms of the common experience they share and the qualifications of the individuals with whom they have trained. Assessment of professional competence based on this approach is simple; it consists merely in documenting that the candidate for licensure or relicensure, certification, or recertification has participated fully in the proper rituals, under the direction of an approved priesthood. In contradistinction is that approach which defines members of a profession in terms not of shared experience but rather in terms of shared competencies. Assessment based on this approach requires that candidates demonstrate that they can complete specified tasks at a satisfactory level of performance; information about how they learned to do so, the rituals they endured or the priests they served is totally irrelevant.

The Status of Regulations Governing Licensure and Certification

For purposes of initial licensure or certification, governmental authorities and professional associations have been a bit schizophrenic in that they have frequently tried to have it both ways. Typically they have used the first approach--documentation of training experience--as a means of determining eligibility for admission to an examination that purports to be based on the second approach--the documentation of what the candidate knows and can do. Given this history and what seems to me to be the totally compelling logic in favor of the second approach it is astonishing to find that the public, the regulatory agencies and the professions themselves have, almost without exception, embraced further training rather than acceptable performance as evidence of continued competence.

In response to public demands for assurance regarding the competence of professionals an ever expanding group of states is requiring relicensure of practitioners in an increasing number of professions. These state legislatures are overwhelmingly choosing to require continuing education as a pre-requisite to relicensure. Right now, mandatory continuing education designed to keep professionals abreast of new knowledge is required in various states for eleven professions, and the rate at which additional states are initiating such requirements indicates that the mandatory movement will likely include all the licensed professions within a few years.

Much of this legislation relates to the health care field. For example, over half the states now require, and most of the others are considering, mandatory continuing education for renewal of physicians' licenses. Over half of the states already have or are now considering mandatory CE for dentists. Practically all states have already implemented such a requirement for optometrists and most have also done so for nursing home administrators. Over a third of the states have enacted such requirements for veterinarians; about a dozen states have done so for pharmacists, and almost as many have introduced them for nurses.

But it is not only the health professions which are being regulated in this manner. Lawyers in 12 states are now required by law to participate in continuing education activities, and such legislation is pending or under consideration in over a dozen more states. As of April, 1977, 11 states had enacted and another 26 were considering, mandatory CE for relicensure of realtors. At that time approximately half the states had already passed similar legislation governing certified public accountants and 6 of the 19 states that regulate social work had mandated continuing education for social workers. (Watkins, 1977)

In addition to the state requirements various professional associations have established requirements of their own for continued membership. For example, practically all state medical societies now require continuing education for continued membership. All twenty-two constituent boards of the American Board of Medical Specialties have now endorsed the policy of periodic recertification and several boards have already established the dates and methods that will be used. It is true that with four exceptions this recertification is still voluntary; however, as membership in a specialty association becomes increasingly necessary as a prerequisite for membership in the local medical society, for appointment to hospital staffs and the like, the voluntariness of this policy may be called into question.

While in principle the procedures used in recertification may entail individual assessment, in practice most professional associations, like most state legislatures, have chosen to mandate continuing education rather than evaluation of performance as a basis for continued certification.

Qualitatively, a certificate attesting to the participant's registration for a specified number of CE hours is all that is required, not even attendance at, much less successful completion of, these courses is necessary. Presentation of evidence that the participant has learned something is a totally foreign concept. Quantitatively, the most popular requirement is an average of 25 to 50 hours of CE credit per year and the most popular cycle for recertification is 3 to 5 years.

Limitations of Present Requirements for Competency Assurance

Program Quality

I have estimated that, depending on the precise quantity required and the particular cycle, it would require as much as 25% to 30% expansion in professional training facilities to provide the education mandated in some professions. Can you imagine a 25% to 30% expansion in the number of law, dental, medical or such schools. That is a luxury we do not need and cannot afford. Consequently, continuing education for professionals, which was a minor part of the program of college and university extension divisions a few years ago, is now dominating their curricula today. Insuring the quality of the programs that are resulting from this overnight boom in professional education is a concern of many involved with continuing education. They fear that such programs may go the way of the requirement of continuing education for teachers, described by some as "the worst of any continuing education for the last forty years."

Indeed experience with required teacher education has been so bad that a case questioning its legality was recently brought to the U.S. Supreme Court. In that case the court held that a tenured teacher's rights were not violated when she was fired for refusing to comply with her school board's rule requiring teachers to earn five semester hours of college credits every three years. In contrast with the trial court and the appeals court, both of which held that the school board's action in firing the teacher was arbitrary, the Supreme Court found that the school board's rule "is endowed with the presumption of legislative validity." (Brief Reports, March, 1979)

Regrettably, there is room to doubt the court's "presumption of validity", on two grounds: program quality and program efficacy. Reservations about program quality appear to be more than justified in view of widespread practices with respect to program evaluation. Experience in the legal profession is not untypical. In the February issue of its Brief Reports Educational Testing Service reports that "to date no attempt has been made to evaluate the quality of continuing legal education (CLE) programs or to provide information that can serve as a basis for improving the CLE field." This, despite the fact that lawyers have for years been relying on CLE provided by various sponsors to update their knowledge, and state mandated requirements have been in existence for over three years, and the amount of resources invested in this effort (an estimated 45,000 persons participating in 300-400 courses during the month of February alone) is stupendous and increasing!

Program Efficacy

But even the most excellent program may be wasteful because it is inappropriate. The view that education is the panacea for maintaining competence is based on the assumption that knowledge deficits, especially those concerned with new developments, are the major causes of inadequate performance on the part of practicing professionals.

Given the fact that a professional career may span 40 to 50 years during which the practitioner may be subjected to rapidly accelerating rates of change in the scientific and technological context of practice this is a matter for real concern.

It was my nephew, who at age 3, made me poignantly aware of the magnitude of these changes to which our nervous systems are in fact required to adapt. My great enlightenment occurred on a nice, sunny afternoon when I thought Mike should go outside and play. Instead he was sitting totally committed to a cowboy movie on television. In a misguided effort to inspire him to move I said, "Mike you know when I was a little girl we didn't even have television." Getting carried away with the memory of such deprivations I went on to explain that when I was a little girl it took two days to get to grandmother's house by train because there wasn't a plane to take us there, and on I went. Finally, Mike, feeling he had tolerated this harangue long enough, looked me straight in the eye and said firmly, "Aunt Mickey when you were born had fire been invented?"

Now I am quick to insist that fire had been harnessed and already existed as a useful tool when I was born, but I am also ready to admit that during my professional lifetime critical changes have occurred that threaten to render many of us obsolete. At the risk of having you share my nephew's view of the relation between the date of my birth and that of the birth of fire may I draw your attention to only one such change in our own field. When I was a newborn professional not only were there no mini-computers, there were also no maxi-computers, no remote terminals and no inexpensive little calculators with or without sophisticated functions. If you wanted to compute a correlation--simple, partial or multiple--you either did it by hand or, if you were lucky, you hired a research assistant to do it by hand or, if the gods really smiled on you, you or your research assistant obtained access to the one monster calculator in the department which, because it could add, subtract, multiply and divide electronically, was kept locked up in the Chairman's office. Think what this single change has done to or for research design in our field! Consider its implications for assuring and assessing the professional competence of educational researchers whose professional birth predates computers.

But if change presents a problem in our own relatively slow-moving field consider the situation in those professions which are more closely aligned with the rapidly moving biological and physical sciences. For example, some physicians are fond of reminding us that it was not too long ago when the most reliable laboratory procedure for diagnosing diabetes was the urine-taste test. Nowadays, with a single check on a laboratory order blank, physicians regularly learn much more than they care to know, and maybe more than they can readily interpret, about the composition and possible contamination of any body fluid.

The implications for the process of medical problem solving of this revolution in the technology and economics of laboratory testing were dramatized at a recent workshop I conducted in Bangkok where internists

from Uganda, India, Thailand, Sri Lanka and the Philippines debated with their colleagues from Australia, Britain, and the United States the optimal approach to the workup and management of patients presenting with common symptoms that are to be found everywhere in the world. The former group stressed the importance of ordering very specific tests--a hemoglobin or hematocrit--the latter group routinely argued for a complete blood count, urinalysis or even SMA 12 or 24. Not only the technology but the economics of the health care delivery system places quite different demands on the physicians practicing in these different settings and thus alters the relative importance of different components of competence.

The problems of maintaining professional competence are not those of simple "future shock". Rather these problems can be viewed as a consequence of differential rates of change. Changes in the technological base of our society are occurring at logarithmic rates; institutions are responding at arithmetic, or at best geometric, rates; meanwhile our individual genetic makeup is changing so slowly and so imperceptibly that our nervous systems may simply be unable to accommodate to the new conditions.

Under these circumstances how do we assure continuing professional competence? Is mandatory CE the answer? I doubt it. There seems to be a mounting body of evidence to the contrary. For example, Bashook (1979) reports that in over 200 medical audits, sampling some 125 diagnostic categories in practically every medical specialty, not one instance has been found where failure to meet criterion was attributable to an educational need that could be remedied by a formal educational program. True these audits were carried out in a tertiary care center where an occasional individual might require some very specific information. But most deficits were correctable by individual discussion or by organizational and/or physical changes. (i.e. making a portable X-ray more conveniently available, having certain types of personnel available at off hours, etc., clarifying responsibilities of different personnel, etc.)

Williamson (1976) reports that in studies of performance of medical laboratories it has been found that most error rates increase significantly in many laboratories after 5 p.m. You may ask what does the clock have to do with it. Well, 5 p.m. is when the regular staff leaves, the auto-analyzers are turned off, and older assay methods are employed. Even more significant, proficiency testing in laboratories reveals no relation between professional performance and formal certification status. Board certified pathologists, for example, apparently do no better than those who are not certified. These findings not only raise serious questions about the validity of laboratory reports under usual conditions, and suggest that intensive investigation of the whole subject may be required, but also suggest that situational constraints and pressures may be at least as significant as educational opportunities and limitations in affecting professional performance.

For example, one of the most impressive changes in staff performance is documented in the famous 1956 Lembcke Study which reported that in one

hospital the number of unnecessary pelvic operations was reduced by about 90% after introduction of repeated medical audits of records involving all pelvic surgery. In this case, a summary report was made to the appropriate hospital committee, following which the "president of the medical staff, assisted by physician members of the joint liaison committee, interviewed the surgeons with poor individual records and sought their compliance," after which the criteria were distributed to staff. Since no other informational or educational interventions occurred, this study has been widely interpreted "as illustrating the impact of a unified, genuine commitment of the staff, administration and board to improve the quality of care, once the precise deficiency and those responsible for it have been objectively and systematically identified by audits based on essential criteria." (Sanazaro, 1976)

Some of the best evidence supporting education as a means of assuring continuing competence is to be found in Payne's (1972) excellent study of medical care in Hawaii. He observed that while statistically significant increases in the physician performance index were registered for most problems following very specifically targeted educational programs, performance was still well below criterion. But in these studies as in others, analysis of the reasons for failure to meet standards rarely revealed lack of knowledge as the primary, or even a major contributing, cause. Unavailability or inconvenient location of equipment, shortage of personnel, inadequate institutional organization and requirements, and low institutional commitment to criteria seemed to be of over-riding importance.

Alternatives to Present Practice--The Issues

Given these serious reservations about mandated education as a means of assuring continued professional competence, other alternatives must be considered. Fortunately, some of the newer professions and occupations furnish some guidance in this regard. For example, I am informed (reliably, I hope) that persons who wish to maintain a valid commercial pilot's license must present themselves every six months for what is supposed to be a thorough physical examination, a comprehensive psychological evaluation and a searching performance test, the latter conducted on both a simulator and a real aircraft of the type for which license is sought. Clearly this suggests a totally different approach to assuring professional competence and raises a fascinating set of issues that must be dealt with when capacities and skills rather than training are to be documented. For convenience I have grouped these issues as follows:

1. For what professions and for what individuals within each profession should repeated documentation of performance be required? Who needs to be reassessed?
2. How often should individuals be required to present evidence of competence? What is the optimal cycle for reassessment?
3. On what range of competencies or types of competencies do reassessments need to be made for practicing professionals?

4. What techniques are available for making these assessments?
5. What criteria should be used in judging performance?
6. How should standards for continued licensure or certification be set?

Who Needs To Be Reassessed

This is an interesting question and one to which the answer is not entirely obvious. First, I have serious reservations about issuing an unlimited lifetime license or certificate in any professional field. But if a license or certificate is to be limited in any way then some mechanism for renewal must be made available and if the initial license or certificate was issued on the basis of an assessment of competence then, logically, renewal of the license or certificate should entail some evidence of demonstrated competence. Thus, my tentative answer to the first question is that assessment of continuing competence is required for members of all groups for whom the state and/or voluntary associations take any initial responsibility.

But within each of these groups does everybody need to be reassessed at the same specified intervals? Most regulatory bodies and professional associations have responded with a resounding "Yes" to this question, thus requiring that significant proportions of the total membership of the profession be reassessed annually. As we have seen, in the case of commercial pilots twice the membership must be reassessed annually; in the case of the older professions one-fifth to one-third of the membership is more typical. Now, if done properly, reassessing even these smaller numbers would require the investment of enormous quantities of resources that I am not sure we can afford to allocate to this purpose.

Thus it may be necessary to set priorities with respect to the categories of practitioners within each profession who should be reassessed at specified intervals. To facilitate this process it will be important to find a set of indicators in each profession that will automatically identify high risk persons who should be reassessed at relatively shorter intervals. For example, research in the health professions seems to indicate that the setting in which the professional works is associated with the maintenance of competence and that professionals practicing in academic or other group settings maintain competence at a higher level over a longer period than do those who are geographically or psychologically more remote from such group settings. Similarly the social risk or cost of incompetence may be relevant in making a determination regarding the urgency of reassessing certain categories of practitioners within a profession. There may be some professional situations in which the risks of incompetence are so negligible that the priority for reassessing professionals working in those situations is low. For example, among pilots those who fly a Cessna 172 are potentially a lesser social menace than are those who fly Boeing 747's. In short, a requirement that every professional be reassessed at relatively short intervals is likely to result in such heavy social burdens and such

high social costs that the system will fail. It is therefore imperative to establish priorities in terms of cost-benefit considerations that are realistic and can be implemented.

The Optimal Cycle for Reassessment

In making a decision about the appropriate cycle for reassessment it is essential to consider the nature of the competencies required in the field and the likely rate of decay in them. In a profession in which knowledge and practice are rapidly outdated or where skills become quickly dulled by bad habits or inadequate practice, frequent reassessments at short intervals are clearly indicated. In other professions where basic knowledge and skills once acquired are not likely either to be lost or to become outdated the appropriate interval for reassessment should be quite lengthy. Indeed, there may be some professions, possibly social work, for example in which the most significant changes are those concerned with the availability of social resources and the procedures for obtaining their use. Reassessment in such fields would not need to occur on a regular cycle at all but could be made to depend on ad hoc decisions in response to significant external social changes. Thus, despite fads to the contrary, there is no universally optimal cycle; this issue must be independently resolved for a profession in light of the nature of the essential knowledge and skill that defines each. This leads us to consideration of the next issue; namely, what competencies should practicing professionals be required to demonstrate?

What Competencies Should Practicing Professionals Be Required to Demonstrate?

It has become a cliché in the health professions to quote evidence which seems to indicate that the practicing professional reaches a peak of competence some 5-10 years after completion of formal training and goes rapidly downhill thereafter. Now the interesting thing about these data is that to the best of my knowledge they have been collected by using test instruments that are employed for purposes of initial licensure or certification. Typically this instrumentation places heavy emphasis on conventional multiple choice tests designed to survey a broad universe of knowledge. My hypothesis is that people do not begin their professional decline quite so early but rather that the apparent decline is a function of increasing specialization manifested in the form of declining scores on instruments designed to sample general knowledge.

Whatever its cause the phenomenon has serious implications for assessing practicing professionals: Should the lawyer who has spent the past 5 years as trial lawyer in criminal cases be required to demonstrate the same competencies as one who has spent the same period in the board rooms of a single corporation, or as one who has been buried in the offices of a federal regulatory agency?

There is no easy answer to this question. Some groups have argued that an individual profile be constructed for each professional which indicates the nature of his/her practice, and that any reassessment

should be made only in terms of that practice profile. Others have suggested a formal audit of the individual's practice such that only those competencies directly involved in the day-to-day practice of that individual would come under scrutiny. Still others have recommended that assessments of continuing competence be limited to, or heavily weighted toward, knowledge of new developments in the field. Thus, for example, an individual applying for relicensure or recertification would not be assessed on basic, common, general knowledge and skills but only on those where significant changes had occurred since that person's last assessment.

While there is a certain compelling logic to these views they fail to account for the repeated finding that failure to act in accord with professional standards is very often a function of bad habits and/or inappropriate responses to various types of constraints. If so, then assessment of continuing competence cannot be limited to knowledge of new developments and/or narrow technical skills, but will need to encompass those component attributes which each profession defines as essential to effective performance. The only alternative which I can see is limited licensure which recognizes specialization.

What Techniques Are Available For Assessing Competence of Practitioners?

Whether it will lead to renewal of a limited or of a general license, a determination regarding continuing competence should, in my view, be based on an assessment of performance.

Generically, I believe there is only one technique for making such a determination and that is to observe an actual performance of the relevant kind. However, it is important to recognize that the observed performance may occur in a "natural" or in a "contrived," situation. I have deliberately used the word "contrived" rather than "artificial" or "test" situation, because it seems to me that these other terms are misleading in suggesting that performance in one situation is "real" and in the other is, I suppose, "unreal". It is critical for us to recognize that, whether we are observing an individual in a situation that arises naturally in the context of professional practice or in a situation that has been contrived for assessment purposes, the individual's performance in both situations is "real" in every sense of the word. Only the constraints, and the degree of reproducibility, of the two types of situations differ. While these differences have significant implications for the specific inferences we can make they should not lead us to regard one performance as more authentic than the other.

A "Natural" or a "Contrived" Situation

For example, consider the following assessment made by the Center for Disease Control which has responsibility for licensing and proficiency testing of medical laboratories that provide interstate service and all others who voluntarily join the proficiency testing program. To assess participating laboratories blood samples into which a known drug had been introduced (for example, narcotics, barbiturates, or amphetamines) were prepared. Each laboratory was sent half of a split

specimen directly with the request that it be assayed for unknown drug abuse substances as part of the proficiency testing program; assay results correctly identified the substance in 96% of the specimens. The remaining half of each sample was sent to a physician in the community who put a patient's name on it, and then sent it in for assay to the same laboratories with the same request, but without identifying it as part of the proficiency testing program. The accuracy rate in assays of these specimens dropped from 96% to 50%. (Williamson, 1976) Since the situation was in both cases a "contrived" one, was either performance "real", if so, which?

The Reliability of Judgment in the "Natural" Situation

Evaluation of professional performance in the "natural" situation implies judgments about quality and this normally entails some form of peer review. The medical profession has probably had more experience than any other profession in this form of assessment and it is instructive to look at their results. First let us consider limitations on the reliability of peer judgments. One of the earliest studies of this subject was undertaken in 1934 by the American Child Health Association. In that study a sample of 1,000 11-year-old children was utilized to determine concordance of physician judgment regarding indications for tonsillectomy (T&A). On examination it was found that 611 of the sample had already had a T&A. The first of three independent panels of physicians examined the remaining 389 children and judged that one in three required a tonsillectomy. Those judged by the first review panel as not requiring a T&A were then examined by a second panel which oddly enough recommended T&A for approximately one-third of them. The 116 remaining at this point were then examined by a third panel which (predictably?) judged that roughly one-third of this group required a tonsillectomy. Overall, in only 65 cases among the initial sample of 1,000 was a T&A judged to be not indicated. The investigators commented that it was fortunate for those remaining few that the study had run out of physician teams to provide further independent examinations. (American Child Health Association, 1934)

The results of more recent global reviews of office or hospital records are instructive. For example, in a retrospective review designed to appraise the quality of obstetric and newborn care in 1,258 cases involving major obstetric complications, recorded in the Rochester Hospital Region, it was found that in 33% of the cases, one of the three expert reviewers disagreed with the other two in categorizing care as "satisfactory" or "unsatisfactory". (Richardson, 1972) In subsequent studies conducted by the same authors checklists designed to serve as guides to the ultimate decision of "satisfactory" or "unsatisfactory" were developed. Even with full utilization of these checklists it was found that judges differed so greatly in the harshness of their judgments and/or in the relative weights each assigned to different aspects of clinical care that, on the average, it would have required at least 16, and in some areas as many as 28, expert judges to achieve a stable composite judgment--at the 95% level of confidence--that the care rendered in individual cases was or was not "satisfactory". (Richardson, 1972)

In response to these difficulties the medical profession has, over the past 5 to 10 years, experimented with a variety of audit procedures that yield more readily reproducible results but as we shall see when we consider audit criteria, obtaining reproducible results has not eliminated some fundamental problems in assessing performance in vivo.

The Problem of Sampling "Natural" Situations

In utilizing "natural" situations or even "contrived" situations that imitate nature closely, it is important to consider the conditions under which a sample of professional behavior is to be obtained. Some of the studies of the extent and causes of interpretive errors bear on this issue.

In one of the most thoroughly validated such study, replicated repeatedly both in Europe and in this country, serious lesions were missed in roughly 25 percent of X-ray films from 100 randomly selected tuberculosis patients. No relation between the number of errors and the size or character of the lesion could be established. (Williamson, 1976) Investigators concluded that there seem to be limits to human performance in reading such films under the conditions and pressures of average practice and that to avoid such errors it may be necessary to require dual readings, or to have radiologists read smaller numbers of films at a single sitting. By the same token it may be necessary for those who assess the competence of radiologists to consider carefully the conditions under which they will sample performance.

Similar results have been obtained in analogous studies of cardiologists' interpretations of electrocardiograms. In one study for example, there was major disagreement among independent readings in 20 percent of the tracings from 100 acute coronary suspects. Of equal concern is the fact that physicians disagreed with their own previous readings in 13 percent of tracings fed back to them two weeks later (Davis, 1958), a finding which raises questions not only about the conditions under which performance is to be sampled but also about the size of the performance sample to be obtained.

In this context it should be noted that the generally accepted position is that assessment of competence should be primarily concerned with patterns of practice rather than single incidents. That is, for any individual professional, an unexpected catastrophe can be completely inconsistent with the usual pattern of practice, and it is only repeated instances of inadequate performance that should concern the licensure or certification body. Parenthetically, it must be noted that this attitude is antithetical to that of legal experts who tend to cite individual exceptions rather than patterns of performance. (Task Force, 1974)

Determining What Criteria Shall Be Used For Judging Performance

Criteria for judging the quality of professional performance in either natural or contrived settings may be either implicit or explicit, and they may involve only process, i.e. what the professional did, -or they may involve only outcome, i.e. the end result of the process, -or

they may involve both process and outcome.

In an interesting study comparing the effects of using these different types of criteria on the same 296 patient records, Brooks and Appel found that when explicit, process criteria were used only 1.4% of the cases were judged to have met acceptable standards, whereas 63.2% of the same cases were judged to have met acceptable standards when implicit, outcome criteria were used. (Brooks and Appel, 1973)

But the issue of "process" vs. "outcome" is not the only problem area in the determination of criteria. Some would argue that the real issue is whether the criteria should reflect essential or optimal actions, i.e. whether they should include only those actions or procedures essential in every situation of a specified type, or should include everything that might possibly be involved in the proper handling of that type of situation. In a recent study of this issue Studnicki and Saywell (In Press) applied two different audit methods to 6,980 patient records in eight diagnostic categories, within 22 Maryland and Pennsylvania hospitals, representing the performance of 1,321 attending physicians. One of the audit methods they used yields a Physician Performance Index (PPI) which represents the proportion of recommended procedures actually supplied by the physician. According to this method a perfect score is 100%. The other audit method which they employed flags unjustified variations (Flags) from standard patterns of behavior. According to this method a perfect score is 0% flags. Thus, ideally one should expect to find a perfect inverse relationship (i.e. a correlation of -1.0) between results of these two audit methods. However, the authors actually found correlations ranging from a high of -.88 in one diagnostic category to a surprising positive correlation of .14 in another.

We seem to be left with an insoluble dilemma: On the one hand, data such as those quoted above rightly suggest that adequate linkage between process and outcome has not been adequately established in most fields. On the other hand both for individuals and small groups the results of outcome assessments are exceedingly difficult to interpret. Indeed, the person who is most highly regarded by colleagues may actually look worst in terms of outcome assessments simply because that individual's practice contains an abnormally high proportion of the most complicated problems.

Considerations such as these may lead us to consider utilizing more "contrived" situations where the criteria may be more widely shared. In this connection I would refer you to experience with the Objective Clinical Structured Examination (Harden and Gleeson, 1979), which is designed to simulate the natural situation in a highly controlled manner.

How Should Standards Be set?

The Level of Acceptable Performance

This brings us to the difficult matter of establishing the level of acceptable performance for the practicing professional. In my view there is simply no issue here; there is only the hard task of applying the theory of criterion referenced testing to a broad spectrum of samples

of professional behavior obtained in a variety of types of settings. The simple question in each case is: how many and which departures from criterion are sufficiently serious to render a person "incompetent"? This question must be answered whether we are dealing with performance in a natural or contrived setting; and it must be answered in each of those types of settings irrespective of the type of competence being assessed.

While implementation of a criterion-referenced approach may present many practical difficulties, in my view it presents no theoretical problems; and, if I may say dogmatically, it is the only proper approach to setting standards for assessment of continued professional competence.

Balancing Different Kinds of Standards

The one other matter that has to be faced in determining the criteria for continued competence is the matter of deciding the extent to which superb performance on certain tasks may be used to offset marginal or clearly unacceptable performance on other tasks. For example, suppose a physician seeking recertification as a thoracic surgeon is able to provide evidence of superb technical skills but fails to meet criterion on mastery of new knowledge. Can superior performance in the one area make up for deficits in the other? At this point I would recommend a departure from pure theory. Specifically, I would recommend that for each component of competence to be assessed, the responsible agency should develop not one-but two--standards, one which everyone agrees represents clearly satisfactory performance and a second which everyone agrees defines clearly unsatisfactory performance. The term "marginal" or "contingent" or some other euphemism may be used to label the band separating these two pre-determined standards. The responsible agency must then develop rules of the game to determine the disposition of those candidates who fail to perform at a "clearly satisfactory level" in all components of the assessment process. My own inclination would be to withhold certification entirely from anyone in the unsatisfactory zone on any performance component and to issue only a temporary or limited certificate to anyone who fell in the marginal zone on any component, while providing this latter group with instructions for bringing themselves into full compliance as promptly as possible.

Conclusion

Historically, we have taken two approaches to assessment of professional competence: the earliest, now enjoying a resurgence in regulatory agencies, validates competence on the basis of the method of achieving it. The one recommended here insists that competence be validated on the basis of relevant performance. Further it seems clear that the traditional achievement tests so effective in general education and so widely used for initial licensure and certification are totally inappropriate for assessing the continuing competence of the practicing professional. What is needed is an adequate sample of performance in natural situations or in carefully structured "test" situations that closely simulate nature. Though use of such situations raises difficult issues of sampling, observing, interpreting and judging the adequacy of performance, no other approach is viable.

REFERENCES

- American Child Health Association: Physical Defects: A Pathway to Correction. New York, Am. Child Health Ass'n. 1934. pp.80-96.
- Bashook, Phillip. Personal Communication. 1979.
- Brooks, Robert H. and Appel, Francis A. Quality-of-care assessment: choosing a method for peer review. The New England Journal of Medicine. Vol. 288, No. 25. pp. 1323-1329. June 21, 1973.
- Davies, L.G.: Observer variation in reports on electrocardiograms. British Heart Journal, 20:153-161, April 1958.
- Educational Testing Service. Brief Reports, February-March, 1979.
- Harden, R.M. and Gleeson, F.A. Assessment of clinical competence using an objective structured clinical examination (OSCE). Medical Education Booklet No. 8. Medical Education, 13:41-54. 1979.
- Institute of Medicine. Report of a Study Assessing Quality in Health Care: An Evaluation. National Academy of Sciences. Washington D.C., 1976.
- Matarazzo, Joseph D., Higher education, professional accreditation and licensure. American Psychologist. Vol. 32, No. 10, pp. 856-7. October, 1977.
- Payne, B.C. and Lyons, T.F. The Methods of Evaluating and Improving Personal Medical Care Quality: Vol I and II. University of Michigan, Ann Arbor, Michigan, 1972.
- Richardson, Fred MacD. Peer review of medical care. Medical Care. Vol. X, No. 1, pp. 29-39. Jan-Feb. 1972.
- Sanazaro, P.J. Medical audit, continuing medical education and quality assurance. Western Journal of Medicine. 125:241-52. September, 1976.
- Studnicki, James and Saywell, Robert M. Comparing medical audits: Correlation, scaling and sensitivity. Journal of Medical Education. (In Press)
- Task Force on Guidelines of Care: American Advisory Committee on PSRO. PSRO's and norms of care. Journal of the American Medical Association. Vol. 229, No. 2, pp. 166-171. July 8, 1974.
- Watkins, Beverly T. Certification of professionals: a bonanza for extension programs. The Chronicle of Higher Education. April 11, 1977.
- Williamson, John W. The product of our medical schools in perspective. Academic Decision Making: Issues and Evidence pp. 73-96. Washington. Ass'n of Am Med Colleges. 1976.